



Heriot-Watt University
Research Gateway

Big data and ergonomics methods

Citation for published version:

Walker, G & Strathie, A 2016, 'Big data and ergonomics methods: A new paradigm for tackling strategic transport safety risks', *Applied Ergonomics*, vol. 53, no. Part B, pp. 298–311.
<https://doi.org/10.1016/j.apergo.2015.09.008>

Digital Object Identifier (DOI):

[10.1016/j.apergo.2015.09.008](https://doi.org/10.1016/j.apergo.2015.09.008)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Peer reviewed version

Published In:

Applied Ergonomics

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Special Issue: Big Data in Transportation and Traffic Engineering

Using Big Data from Black Boxes to Tackle Strategic Human Factors Risks

Dr Guy Walker & Dr Ailsa Strathie

Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh, EH14 8AS

Corresponding author: Dr Guy Walker, G.H.Walker@hw.ac.uk

Abstract

Big-data collected from On-Train Data Recorders (OTDR) could be used to tackle the most important strategic risk issues currently faced by rail operators and authorities worldwide. These risk issues are increasingly orientated around Human Factors. We prove the concept that long standing Human Factors methods can be driven from big-data, moreover, that their outputs speak directly towards improving the accuracy of future predictions. Over 300 Human Factors methods were reviewed and a smaller sub-set selected for proof-of-concept development using real on-train recorder data. From this are derived nine candidate Human Factors leading indicators which map on to all of the psychological precursors of the identified strategic risks. The intersection of psychological knowledge, Human Factors methods and big-data creates an important new framework for driving new insights and increasing the accuracy of predictions.

Keywords: Big data, leading indicators, human factors

Introduction

Data recording is the act of automatically logging information on system parameters over time. Data recording has become increasingly ubiquitous in rail transport operations and easily qualifies as a 'big data' problem (e.g. Wu & Liu, 2014; Geisler et al., 2011). Entire national train fleets are now required to carry recorders which continuously extract data on how individual trains are being driven, at increasing rates, and across an increasing range of parameters. The outflow of data is therefore extensive and growing in terms of volume, velocity and variety (Laney, 2001). Perhaps because of this, and the conceptual challenges involved in storing and mining such large quantities, the data is currently underused (Hart, 2003). What could it be used for? In this paper we argue it could be used to tackle the most important strategic risk issues currently faced by rail operators and authorities worldwide.

The Spanish train crash in July 2013 reveals yet again that the interface of humans and machine systems is key (Evans, 2011; EU, 2003; RSSB, 2009). Having closed off numerous other technical/engineering opportunities for accidents to occur what is left is a troubling class of accident which arises despite highly trained and motivated personnel, the presence of robust safety management practices, and state of the art infrastructure. More worryingly, the methods and approaches that have helped us achieve the current high levels of railway safety seem to be less effective in the face of human/system problems such as these. By using big-data as an input to long standing Human Factors methods, however, we can embark on an important research agenda that takes us towards leading indicators of strategic risks of precisely this sort. In this paper four proof-of-concept demonstrators are realised, tested with real-world train data recordings, and presented as a highly novel framework for driving new insights and increasing the accuracy of predictions.

The Black Box Paradox

The intersection of big-data and the rail transport context embodies three paradoxes. Firstly, because so few major rail accidents occur there is no longer enough data to construct reliable forward looking estimates (Evans, 2011). When safety performance data reaches the level of that achieved in the rail sector it instead starts to become characterised by unpredictable periodicities, cycles or discrete events. This is becoming evident in EU rail safety data, with a large scale rail accident (on the scale of the recent Spanish rail crash) occurring on average every six years (EU, 2003). Safety data is 'levelling off' with a persistent class of human/system accident now elevated to the status of a key risk (RSSB, 2009; Stanton & Walker, 2011). Secondly, "there is widespread concern within the industry that the background indicators – rather than the headline grabbing ones – have remained worryingly stable" (Wolmar, 2012). An example of this is UK data on Signals Passed At Danger (SPAD) incidents. As a class of accident/near-accident SPADs have been the focus of sustained attention and research in every decade since the British Transport Commission initiated the first Medical Research Council studies of the 1960's. Considerable progress on SPADs has been made but despite the introduction of the Automatic Warning System (1950s), modern colour light signalling (1960s), improved train braking performance (1970s), and subsequently a new Train Protection and Warning System (1990s/2000s), SPADs still occur (McLeod, Walker & Moray, 2005). Data for the past three years indicates in the region of 20 to 30 SPADs per month, approximately half of which fall into high risk categories. More worryingly there has been comparatively little variation in the overall SPAD rate. For example, the rate for Quarter 4 2012 is the same as Quarter 3, and indeed the same (or very nearly the same) as on seven previous reporting periods since 2005 (e.g. ORR, 2012). The third and final paradox is that the opportunities to use On-Train Data Recorders (OTDRs) for their original purpose (i.e. post-accident analysis) are diminishing at the same time as the technical capabilities of data recorders are increasing (Geisler et al., 2011; Morcom, 1970).

What this means is that Exabyte's of non-accident data are being collected day in and day out, but not currently used. The opportunity embedded in this is to use big data from transport recordings to detect accident precursors, but specifically those accidents which have proven resistant to our current approaches and which are responsible for the current plateauing of accident trends.

Human Factors Leading Indicators

Leading indicators are measurable precursors to major events such as an accident. The indication of a precursor 'leads', or comes before, the actual event itself. Lagging indicators are the opposite. These are so called 'loss metrics' that can only become apparent after an event (Rogers, Evans & Wright, 2009). Leading indicators are said to be 'proactive' because they enable steps to be taken to avoid seriously adverse consequences. Lagging indicators are said to be 'reactive' in that a seriously adverse event needs to occur before it can be learnt from. For this reason, leading indicators are also sometimes referred to as 'positive performance indicators' and lagging indicators as 'negative performance indicators'. The concept of leading and lagging indicators originally derives from the field of economics and the need to understand 'business cycles' and to predict when one phase of a 'cyclical process' such as this will change to another (Mitchell & Burns, 1938). The terms have been appropriated more recently by the safety and risk field, particularly in view of developments in Safety Management Systems (SMS) since the 1990's. Leading indicators, in a Safety Management context, can be defined as "something that provides information that helps the user respond to changing circumstances and take actions to achieve desired outcomes or avoid unwanted outcomes" (Step Change, 2003, p. 3). That 'something' can be defined according to the risk factors underlying the troublesome class of operational accident (or near accident) that is the focus of this paper.

The reason for the emphasis on Human Factors can be seen in the 'broad causes' attributed to recent rail accidents. Out of seven 'broad causes' attributable to European rail accident data, four out of seven, including the top three, involve a prominent Human Factors dimension (Evans, 2011). Expressed in descriptive terms, people in these scenarios either 'get out of sequence', 'lost situational awareness', 'allocated attention incorrectly', 'allowed prior experience to override the correct action' or combinations of all four, as shown in Table 1. A comprehensive review of the psychological literature (reported elsewhere, Walker & Strathie, 2012) linked these descriptive terms to a set of specific and robust risk factors (these are also shown in Table 1). If we are able to detect when these risk factors are present, using big-data as the input, then we should also be able to make progress on key strategic Human Factors risks.

Table 1 – Strategic risk issues, descriptive causes, underlying Human Factors risks and their source in the wider literature

Human Factors accidents	“Getting out of sequence”	“Loss of situational awareness”	“Faulty allocation of attention”	“Prior experience overrides correct action”
Signal Passed At Danger				
Signalling Error				
Over-Speeding				
Other Operational Error				

	Risk Factor	Selected references
“Getting out of sequence”	<ul style="list-style-type: none"> • Task interruptions • Where an unusual response must be made at a future point in time • A device or system that only needs infrequent monitoring • Removal of normal environmental cues that trigger habitual behaviour • When circumstances mean that a driver/air crew must deviate from a well-practiced procedure in a single respect. • Performing monitoring tasks may negatively impact (slow) the main task. 	Graf, 2011; Park et al. 1997; Dismukes, 2007; Meacham & Leiman, 1982; Hicks, Marsh and Cook, 2005; Loft, Smith and Bhaskara, 2011.
“Loss of situational awareness”	<ul style="list-style-type: none"> • Information in the environment which is required to ensure successful task performance is either not present, or not detected. • Failure to understand what the environmental cues received (or not received) might mean. • Failure to accurately project future states (and take appropriate action) based on cues perceived and understanding gained. 	Endsley, 2012; Stanton, Chambers & Piggot, 2001; Gobert, 1998; Salmon et al., 2009.
“Faulty allocation of attention”	<ul style="list-style-type: none"> • When attention is focused on completing a central goal, unexpected events may be more difficult to detect. • When attention is distracted, or is spread across a number of demanding tasks, performance will suffer. 	McMains & Kastner, 2011; Groeger et al., 2005; Neisser, 1979; Simons & Chabris, 1999.
“Prior experience overrides correct action”	<ul style="list-style-type: none"> • Highly routine journeys, particularly where there are long periods during which minimal inputs/responses are required. • Experienced drivers/pilots, for whom the tasks have become part of a procedure that doesn’t require conscious thought. • Systems with more than one mode, where the same responses can produce different outcomes depending on the mode the system is in. 	Bartlett, 1932; Ferris, Sarter & Wickens, 2010; Norman, 1981.

	<ul style="list-style-type: none">• Where the driver/pilot receives conflicting information which results in a situation being misclassified.	
--	---	--

The basic 'research problem' can be stated thus: despite considerable improvements in safety performance in the rail sector, a persistent class of accident/near accident continues to occur. These incidents reside at the interface of people and systems. What is required is a means to detect the presence or emergence of such problems before they manifest themselves as a serious operational accident. This paper describes how big data from OTDRs can be used to 'drive' established Human Factors methods to provide leading indicators of specific risk factors. What are presented are four proof-of-concept demonstrators which show how these approaches work, selected on the basis of their potential scalability when used with big-data.

Methodology

Data File and Parameters

The study uses real-world On-Train Data Recording (OTDR) data. The OTDR data file is a continuous download from a single traction unit. The recording started at 05:34:57 on the 6th July 2012 and ceased at 21:36.32 on the 11th July 2012. This is a period of 136 hours, 1 minute and 35 seconds during which the train made 107 journeys and travelled 1638 miles. The raw data takes the form of a Comma Separated (CSV) file containing a data matrix 191,021 time samples (rows) deep by 72 parameters (columns) wide: a total of 13.8 million data points. The logger itself scans the parameters for changes at a rate of 20mS but, in the present system, to economise on memory requirements data are only logged when one of the 72 parameters changes (up to the maximum scanning/sampling rate). In the present case the mean sampling rate was 2.56 seconds. The OTDR device itself was a UK Railway Group Standards compliant Arrowvale unit which recorded 72 parameters, 25 of which are in addition to those mandated. In terms of data classification four of the parameters; time, distance and two speed signals derived from a driven and non-driven axle, are continuous

ratio data. The remaining 68 are nominal/binomial (i.e. on or off). Explanations for the channels relevant to the present analyses are contained in the proof of concept descriptions.

Rolling Stock

The sample of OTDR data was obtained from a Class 153 'super sprinter', unit number 153 306. This is a single-unit diesel powered railcar built between 1987 and 1988. Class 153s are 23.2 meters in length and have an un-laden weight of 41.2 tons. They seat 72 passengers, comprise a riveted aluminium body shell affixed to a steel under-frame, and are equipped with four electrically powered single-leaf Bode doors. The prime-mover is an under-slung turbocharged 6 cylinder Cummins NT855 diesel engine producing 285bhp. A Cummins-Voith T211r hydraulic transmission drives both axles of the leading BT38 bogie via a Gmeinder final drive. The Unit's maximum operating speed is 75mph. It is fitted with electro-pneumatic clasp brakes, with cast iron brake pads acting directly on the tread of the wheel(s) via compressed air actuation. Air suspension is provided for additional passenger comfort and refinement. Tight-lock compact BSI auto-couplers mean that Class 153's can work flexibly in unison with several other multiple unit classes but the present unit worked solo for the duration of the data collection period.

Journeys and Routes

Data collection took place on the Great Eastern (Route 7) and West Anglia (Route 5) regions of Network Rail. The strategic 'backbone' of the Great Eastern region is the Great Eastern Main Line (GMEL) originating from London Liverpool Street and travelling North East to Norwich. There are numerous branch lines attached to the GMEL providing services to commuter areas such as Upminster, Southend and Colchester, to important freight hubs such as Harwich and Felixstowe, and to more remote communities in East Anglia such as

Sudbury, Cromer and Sheringham. Route 7 (Great Eastern) joins Route 5 (West Anglia) at Haughley Junction, approximately 14 miles from Ipswich, where a secondary route runs West towards Cambridge, Ely and Peterborough. The journey diagram for unit 153 306 between 6th and 11th July 2012 is shown in Figure 1 along with the technical characteristics of those routes in Table 2.

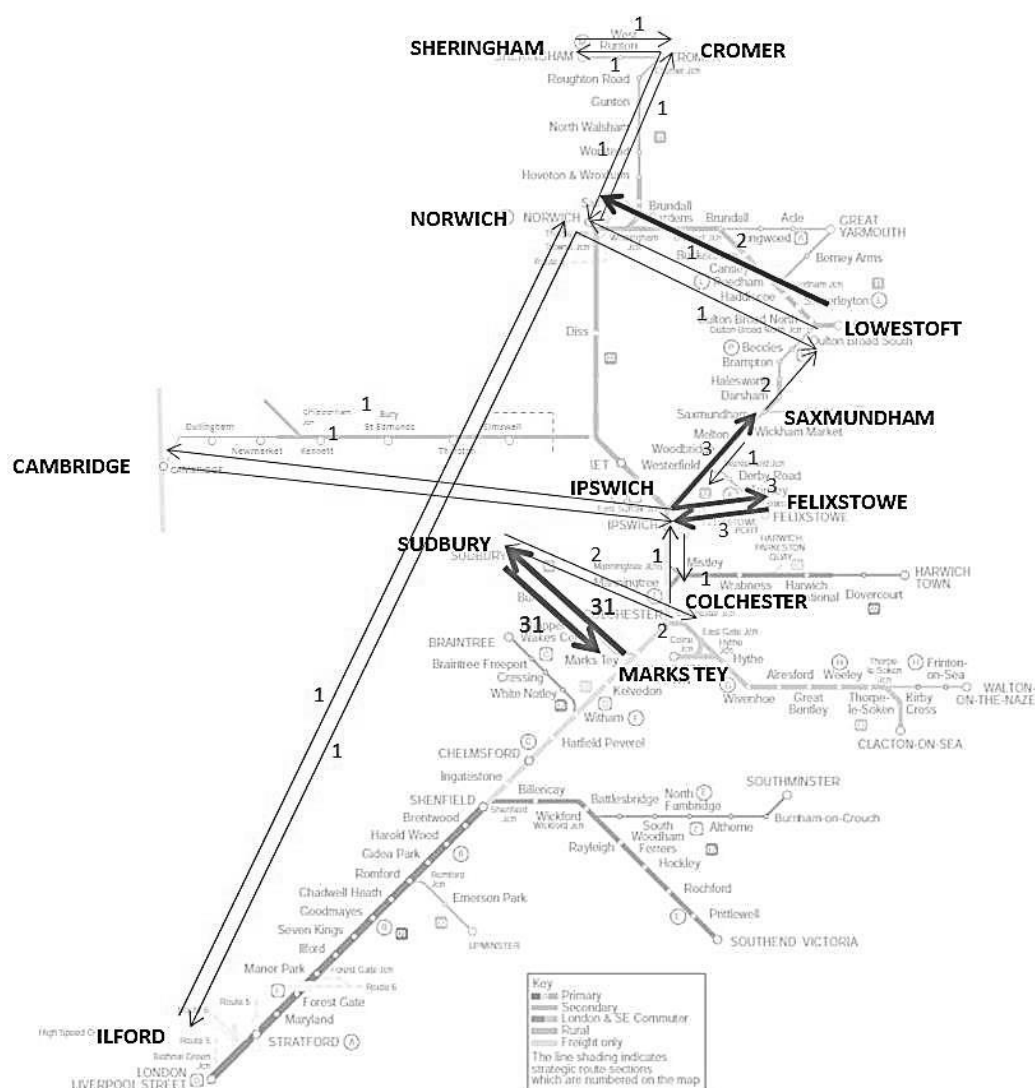


Figure 1 – Journey diagram for unit 153 306 between 6th and 11th July 2012 (excluding local, non-revenue movements of less than 5 miles). Values and line thickness indicate number of journeys made between nodes in the network.

Table 2 - Route characteristics

Route		Class	Freight Gauge	Electrification	Route Availability	Speed	Signalling Type	No of Tracks
Great Eastern Mainline	Route 7	Primary	W10(W9)	25kV AC	8	90(70)	TCB	4(6)
Ipswich - Cambridge	Route 5	Secondary	W10(W9)	None	8	75	TCB(AB)	2(1)
Felixstowe – Ipswich	Route 7	Secondary	W6(W8)	None	7	50	TCB	1(2)
Sudbury – Marks Tey	Route 7	Rural	W6	None	6	50	OTW	1
East Suffolk line and Norfolk Branches	Route 7	Rural	W6(W8)	None	7(6)	Varies	RETB (and other types)	Varies

Freight Gauge

W6 represents 'standard' British Railway dimensions for rail vehicles. W8 compatible routes permit the use of larger/longer shipping container trains. W10 permits 2.9 meter 'Hi Cube' containers and 2.5m 'Euro Containers'.

Route Availability

Principally refers to axle loadings. An RA of 6 equates to a maximum of 20.3 tonnes. An RA of 8 equates to a maximum of 24.1 tonnes.

Signalling Type

Track Circuit Block signalling in which train detection is automatic and train control is (typically) via coloured light lineside signals. OTW = One Train Working, wherein only one train is permitted into a section protected by signalling 'interlocks' until the train leaves the section. RETB = Radio Electronic Token Block, a form of signalling in which a signaller passes an electronic token to the train cab giving permission for it to enter a section of track.

Human Factors Methods

Human Factors (HF) methods provide an explicit way of linking theories on human performance and capabilities to practical situations such as rail operations. Methods are an integral part of the Human Factors discipline (see Stanton et al., 2013 and Karwowski, 2001). The normal inputs to Human Factors methods vary, but common to them all is scale. It is common for methods to be applied to individual scenarios, to accidents/incidents that have already happened, and in rarer cases, to the analysis of entities the size of an organisation. There is no inherent limitation on the scalability of some of the methods, and likewise, no conceptual reason why some methods cannot accept big-data as an input. A systematic methods review was performed to identify candidate methods.

Stage 1 –Review of Existing HF Methods

The result of an initial methods review was a database of over 300 Human Factors methods and techniques, grouped into eleven categories as shown in Table 3.

Table 3- HF method categories

Method category	Brief Description
Data collection techniques	Data collection techniques are used to collect specific data regarding a system or scenario.
Task Analysis techniques	Task analysis techniques are used to represent human performance. They do this by breaking down tasks or scenarios into the required individual task steps, and using these steps to diagnose important human performance/psychological parameters.
Cognitive Task analysis techniques	Cognitive task analysis (CTA) techniques are used to describe and represent the unobservable cognitive aspects of task performance.
Charting techniques	Charting techniques are used to depict graphically a task or process using standardised symbols. The output of charting techniques can be used to understand the different task steps involved in a particular scenario, and also to highlight when each task step should occur and which technological aspect of the system is required.
Human Error Investigation (HEI) and Human Reliability Analysis (HRA) techniques	HEI techniques are used to predict any potential human/operator error that may occur. HRA techniques are used to quantify the probability of error occurrence.

Situation Awareness assessment techniques	Situation Awareness (SA) refers to an operator's knowledge and understanding of the situation that he or she is placed in. SA assessment techniques are used to extract a measure of operator SA while at work within systems.
Mental Workload assessment techniques	Mental workload represents the proportion of operator resources demanded by a task or set of tasks.
Team Performance Analysis techniques	Various facets of team performance can be evaluated, including communication, decision-making, awareness, workload and co-ordination.
Interface Analysis techniques	Interface analysis techniques are used to assess the interface of a system in terms of usability, error, user-satisfaction and layout.

Stage 2 – Initial Methods Screening

Before the HF techniques were subject to further analysis, a screening process was employed to remove any that were not suitable for further consideration due to the following:

- Unavailable – some methods are proprietary. In order to be included, the method should be freely available in the public domain.
- Inapplicable –those methods that did not refer directly or indirectly to the design and analysis of systems, or that were too specialised to be applicable to transport systems, were rejected.
- Duplication – HF methods are sometimes re-iterated and presented in a new format. Any methods that were very similar to other methods already chosen for review were rejected (the version in most common usage and/or most appropriate to transportation environments being selected instead).
- Limited use –quite often a method is developed and not used by anyone other than the developer. Any methods that had not been applied in a practical analysis of some sort were rejected.

As a result of the method screening procedure a list of 87 methods suitable for use within transport systems was created.

Stage 3 – Methods Review

The 87 HF design and evaluation methods were then analysed with respect to the simple question “could the method accept recorder data as an input” and “what output would it provide”. Table 3 presents the list of candidate methods and their possible contribution to the analysis of big-data. The methods met a further three criteria:

1. “can the method be ‘demonstrated’ within reasonable timescales and cost”
2. “constrained by (1) does the method add sufficient value in terms of Human Factors leading indicators”
3. “could the method (at this early stage of analysis) be feasibly automated such that it would require low overhead in terms of external validation, and be compatible with software-based approaches to the management of big-data”

Table 4 – List of candidate methods

Method	Outputs
Workload (Primary Task Measures)	Data recording parameters divided into mental (i.e. warning received etc.) and physical (i.e. control moved etc.) workload categories. Enables crude moment to moment assessment of workload based directly on activity occurring in data recording channels.
Process Charts	Activities, processes, sequences presented in graphical format against common timeline with standardised symbol set. Potentially offers a more ‘user friendly’, higher level, human activity-based representation compared to raw ‘trace plots’.
Link analysis	Nature, frequency and importance of movements/communications between system elements and human operators based on recorder channel activations.
Signal Detection Theory	Response types, categories and timing used to calculate measures of sensitivity and decision bias.

Proofs of Concept

Having linked HF methods to strategic risks on the railway, and revealed the possibility to drive these methods from big-data, this section proceeds to apply them to the OTDR dataset. A brief description of the underlying variables being measured is provided, and the method of application described. Even though the data set comprises in the region of 13 million data points and is, in big-data terms, relatively small, it is clear that useful and interesting outputs were emerging with comparative ease.

Workload Profiles

Brief description

Individuals possess a malleable but ultimately finite attentional resource (e.g. Young & Stanton, 2002). Mental Workload represents the proportion of that resource demanded by a task at any given moment. Excessive demand (over-load) or lack of demand (under-load) can lead to performance problems, in particular, with attention being allocated incorrectly (Theeuwes, 1991). A simple but effective indication of workload can be provided directly from transport data recordings. Certain data recording channels can be associated with mental demand, others with physical demand. As a journey progress a form of workload profile can be derived and analysed to identify the demands being placed on attentional resources.

Method

The proof of concept was tested using a single railway journey from Marks Tey to Sudbury, a distance of 18.5 Km. The 72 parameters extracted from the data recorder were classified into:

1. those that impose a mental demand (e.g. such as an alarm or warning sounding in the cab)
2. a physical demand (e.g. moving the power or brake controllers)
3. both physical and mental demand
4. neither

This gave 14 parameters that relate to mental demand, 20 that relate to physical demand, 6 that relate to both physical and mental demands and 32 that directly relate to neither. Each time one of the parameters changed was recorded in the appropriate category. For the journey selected this gave an overall load of 371 channel activations, 145 (39%) for mental workload and 226 (61%) for physical workload.

Outputs

To create a workload profile, the journey was divided into periods of 10 seconds, and the number of parameters that altered in each was recorded. Figure 2 reveals how mental and physical demands are distributed across a journey, and when workload might be particularly high or low. Overlain on this graph are some further explanatory curves: as the journey progresses from sampling intervals 17-25 how workload declines to a low level. The psychological literature informs us that after approximately 30 minutes of low demand we could expect to observe a vigilance decrement. Depending on the task and context, this can yield a 10/15% reduction in human performance (e.g. Eysenck, 1982; Mackworth, 1948). Time elapsed since workload exceeded a previous value is therefore one indication of potential attention-based risks. The other indication is when a sudden change (as occurs in sampling interval 117) follows a long period of low demand. Again, the psychological literature informs us that attentional resources cannot expand to new demands

instantaneously and that performance problems can arise while they do (e.g. Stanton, Young & McCaulder, 1997).

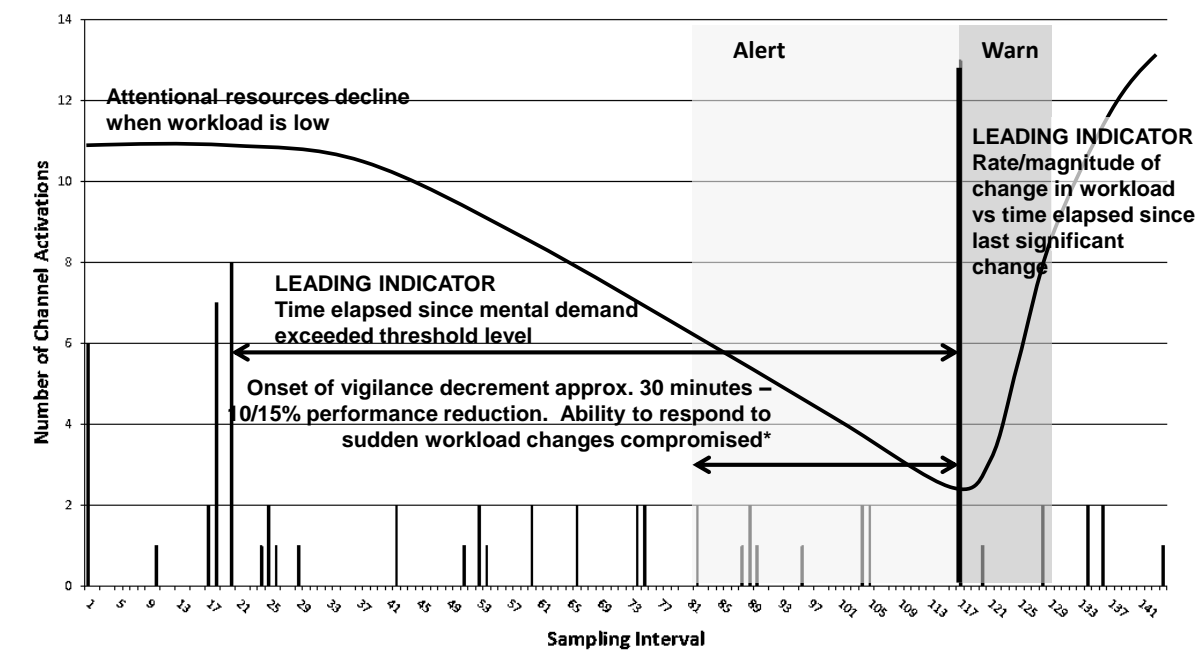


Figure 2 – Recorder channels allocated to mental and physical workload categories and their activation mapped to an actual route.

What are the leading indicators?

Leading Indicator 1: Conventional event detection based on comparing individual samples of workload with mean values across journeys/journey segments/driver population(s) etc. In a wider application, averaging across multiple journeys on the same route, it may be possible to build up a profile of how workload is distributed under typical driving conditions. If changes in the profile appear, this may offer an indication that something has altered in driving conditions that was impacting on drivers.

Leading Indicator 2: Time elapsed since a 'significant' change in workload based on psychological theories of vigilance. While drivers may be more vulnerable to errors during periods of high load, extended periods of low workload can also be problematic. Theoretical knowledge on human performance under the specific conditions pertaining in rail settings could be used to define when occurrences like these constitute early warning of unacceptable risk.

Leading Indicator 3: A 'significant' workload event occurs after a pre-set length of time at low workload has been exceeded. While the example journey analysed here identified periods of high load on arrival and departure from stations (which may not be unexpected), longer journeys on more complex infrastructure may create additional periods of high workload, which may inform practice. The ability to respond appropriately to these changes depends on the magnitude and rate of workload change, and the period of time prior to the change during which low levels of workload were experienced.

Type of Risk Detected: Getting out of sequence because of task interruptions, and/or an unusual response being required. Faulty allocation of attention due to unexpected events, distractors or when attention is spread across a range of demanding tasks. Prior experience overriding correct action due to long periods where minimal inputs/responses are required.

Process Charts

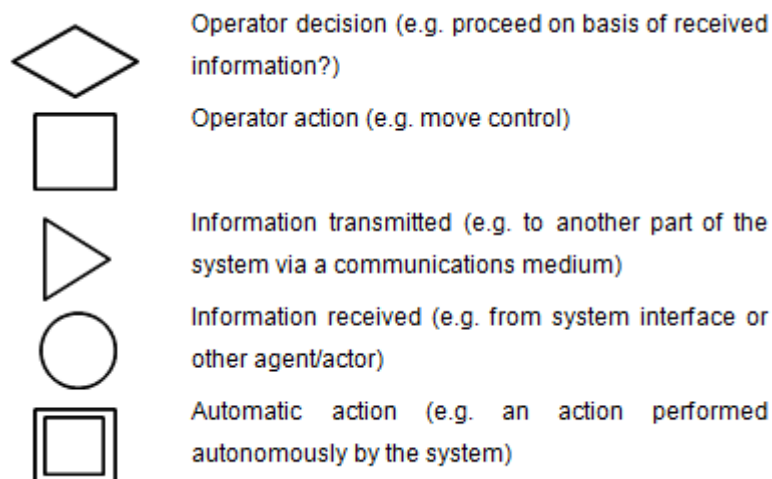
Brief description

Process charting techniques are used to represent complex real-world activity in an easy to read graphical format using standardised symbols and layout (Stanton et al., 2013). The Process Chart methodology has an extremely long legacy and history of use. Early

examples date from the 1920's (Gilbreth & Gilbreth, 1921) and it has been used extensively in military and high hazard domains as a way of understanding the interaction between people and systems, particularly in terms of identifying human error potential. In this application, process charts offer a novel way of converting raw 'trace plots' derived from data recorders into an alternative representation, one that makes it easier to discern how larger journey phases break down into smaller component activities, the order and timing that component activities occur, who is performing what activity and the presence of distinct activity clusters.

Method

The proof of concept was tested using several journey segments spanning a range of different activities and contexts. The 72 parameters extracted from the data recorder were classified into:



Once all of the recorder channels have been classified into the appropriate symbol categories, the process chart itself can be constructed. This involves creating a timeline and columns for each 'agent' in the system. In the case of the railway example six such agents/columns have been used:

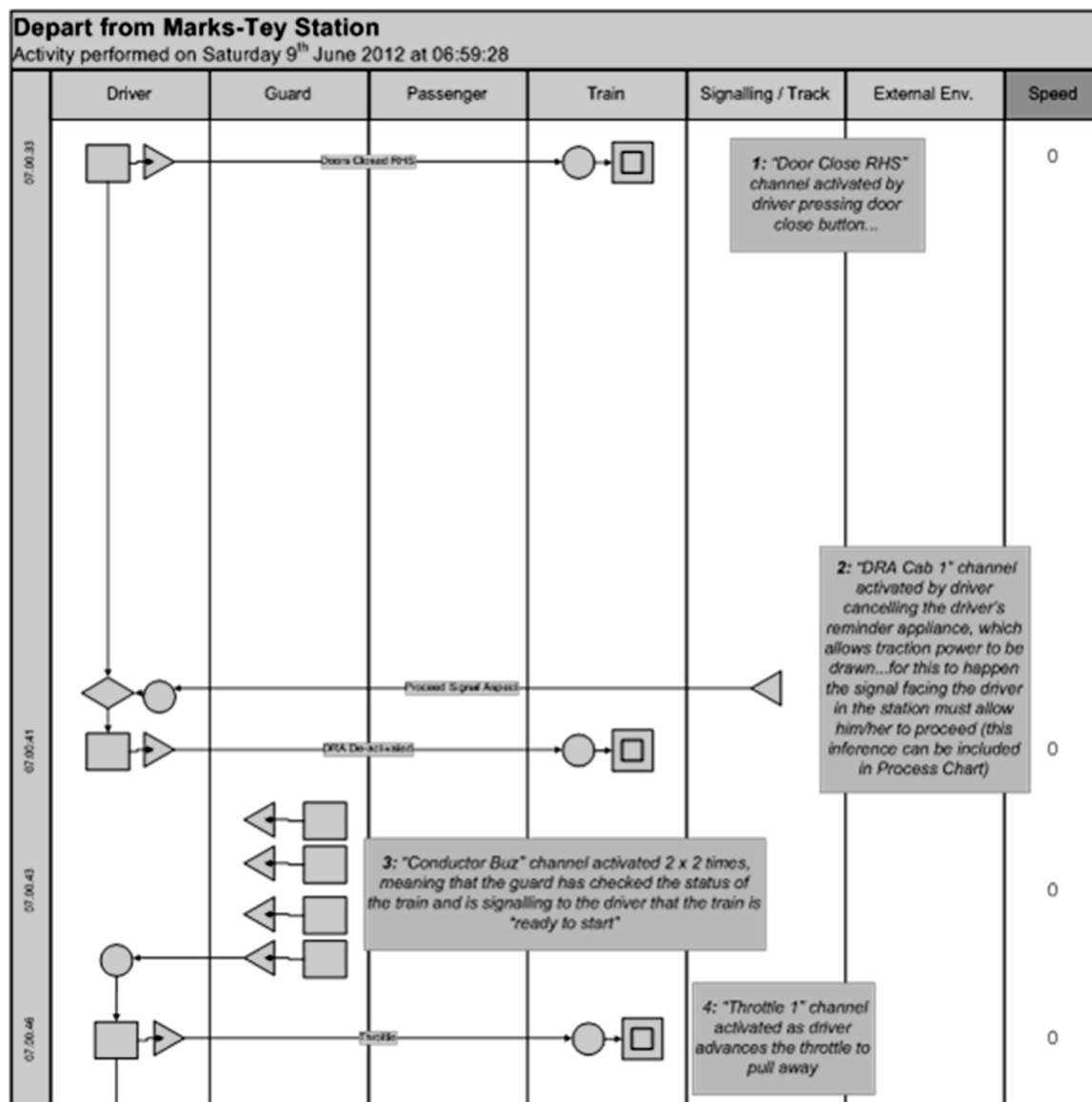
1. Driver
2. Guard
3. Passengers
4. Train
5. Signalling / Track
6. External Environment

As different recorder channels become active, the corresponding process chart symbol is inserted into the relevant column at the correct point on the timeline. The sequence of activities and their dependence on each other defines when these symbols are linked. Thus an activity/symbol that occurs after another activity/symbol becomes linked 'vertically'. Activities that are performed concurrently are linked 'horizontally'.

Output

Figure 3 shows how the channel activations associated with a station departure can be converted into a Process Chart. The boxes have been added to provide a narrative of the activities being performed.

Figure 3 – Process chart depicting activities/actors associated with departing from a station



The simplest output to be derived from a Process Chart, one that links to the Workload Profile above, is an Operations Loading table. This shows how 'busy' each agent in the system is. Whilst the Workload Profile technique provides a high level overview of workload according to Mental and Physical Demand, the Process Chart Method provides a more detailed assessment broken down by different 'actors' in the scenario and specific task/operation types. It is also important to note that the 'agents/actors' represented in the Process Chart are human (i.e. drivers, passengers, guards etc.) and non-human (the train itself, the track and signalling infrastructure and so on). The points at which these different

actors/agents have to interact are clearly shown, along with any potential problems with that interaction in terms of missing information, out-of-sequence and/or overly demanding activities.

Table 5 – Operations loading table

	Driver	Guard	Passenger	Train	Signalling/ Track	External Environment
Decision	1	0	0	0	0	0
Action	3	4	0	3	0	0
Transmit Information	3	4	0	0	1	0
Receive Information	2	0	0	3	0	0
Total Load	9	8	0	6	1	0

The simplifying effect of converting raw trace plots into Process Charts enables meaningful patterns to be better detected and thus made more amenable to automatic detection.

Process Charts also help in the identification of interruptions to the normal sequence of activities and in providing leading indicators around the issue of ‘prospective memory’.

Having the ability to detect behavioural clusters also grants the opportunity to assess whether such structures are typical or atypical. Indeed, whether they are one of a number of different behavioural responses within a wider repertoire, and whether one cluster of behaviour is implicated in risk outcomes more than another, and under what circumstances.

Despite only having access to a comparatively small sample of data distinct patterns of behaviour were still evident. Below are three different ‘clusters’ of behaviour for performing the same task (cancelling an AWS warning by pressing a button when a horn-sound is heard). The first cluster is the normative ‘perceive-decide-act’ sequence. Here the infrastructure on the track triggers an in-cab warning horn. The driver perceives (hears) this, has enough time to classify it (0.89 seconds) and respond by pressing the cancellation button. The second cluster is the ‘predictive cancel’ sequence. In this case the infrastructure on track triggers the in-cab warning horn but the driver responds so quickly

that it is not possible to have perceived, classified and responded to the warning. Instead, the driver has seen the track infrastructure and has anticipated the in-cab warning and timed their response to coincide with it starting. The third cluster is the 'multiple predictive cancel' sequence. As in cluster two, the driver can see the track infrastructure ahead and is pressing the cancellation button numerous times before hearing the in-cab warning horn, and several times after the warning has sounded and been cancelled. Driver behaviour with AWS has been implicated in several high profile rail accidents (e.g. Cullen, 2001; Uff, 2000)

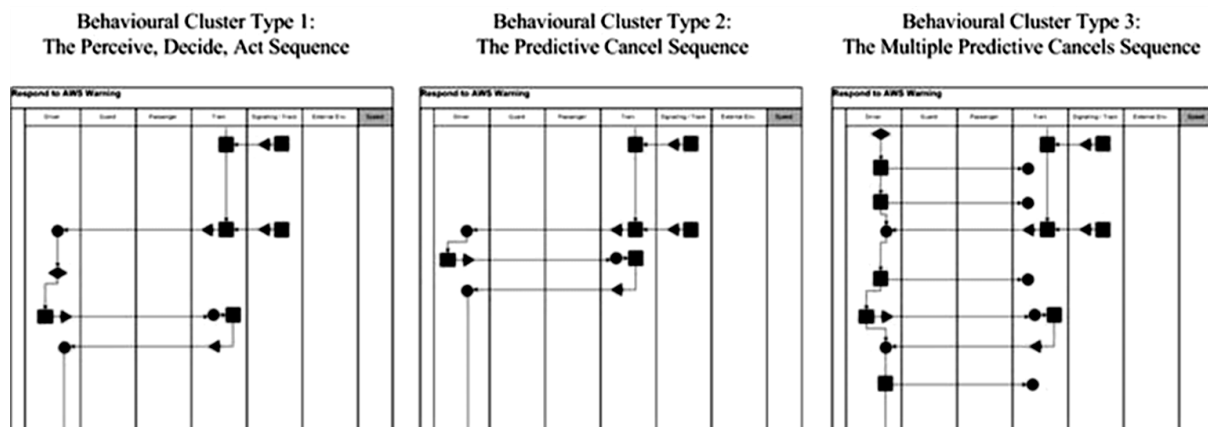


Figure 4 - Three clusters of behaviour associated with cancelling an Automatic Warning System (AWS) alert were detected

What are the leading indicators?

Leading Indicator 4: Operations loading as a means of detecting normal/abnormal workload across different agents/actors in the system. It also reveals the different 'modalities' by which different parts of the system communicate with each other revealing, in turn, sensory channels that might be over or under-relied upon.

Leading Indicator 5: The extraction and definition of behavioural clusters which, in turn, can be explored for their possible impacts on overall risk. Clusters are defined by structure and

sequence of operations. Operations that occur too much or too little, too late or too early, out of sequence or unexpectedly, can be detected.

Type of Risk Detected: Getting out of sequence because of task interruptions, unusual future responses being required, removal of normal environmental cues that trigger habitual behaviour, and deviations from well-practised routines. Loss of situational awareness due to missing information that would normally come from other elements of the system. Faulty allocation of attention due to unexpected events. Prior experience that overrides correct action because of high routine and minimal (new) inputs, changes in system mode meaning that a routine behaviour now causes a different outcome, and conflicting information resulting in situations being misclassified.

Link Analysis

Brief Description

Link analysis is used to identify and represent 'links' between interface components and human operations, and to determine the nature, frequency and importance of these links. Links are defined by a driver's interaction with their cab interface. For example, if someone is required to press button A and then button B in sequence a link between buttons' A and B is recorded. The data matrix created by populating the frequency of these links can be subject to analysis via Graph Theory techniques. Graph Theory is a well-established branch of applied mathematics with a long history of application in different domains. Foundational work in this area occurred well over a century ago and recognisably modern developments occurred from the 1930's onwards. Graph Theory is applied to the analysis of biological, sociological, demographic, economic and computing networks to name just a few. Link Analysis, in which similar principles are applied to 'interface networks' is a more recent development but nonetheless performs particularly well on measures of reliability (e.g.

Stanton & Young, 1999). In this application, link diagrams can be driven directly from recorder data to provide a novel way of detecting events based on people’s physical interactions with technology.

Method

Defining the links between components of a user interface is normally achieved by a walkthrough or observational study of the task(s) under analysis. Recorder data avoids the need for this. Human interactions with cab or cockpit interfaces can be monitored directly based on the changing activation of recorder channels related to cab/cockpit interfaces as shown in Figure 5. For this proof of concept demonstration link analyses performed on six drivers undertaking 17 journeys from Bures to Sudbury.

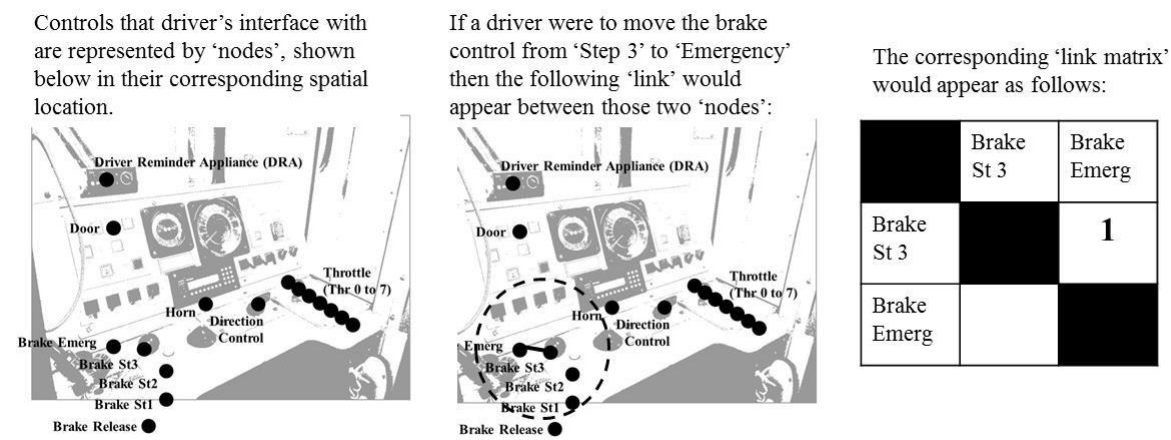
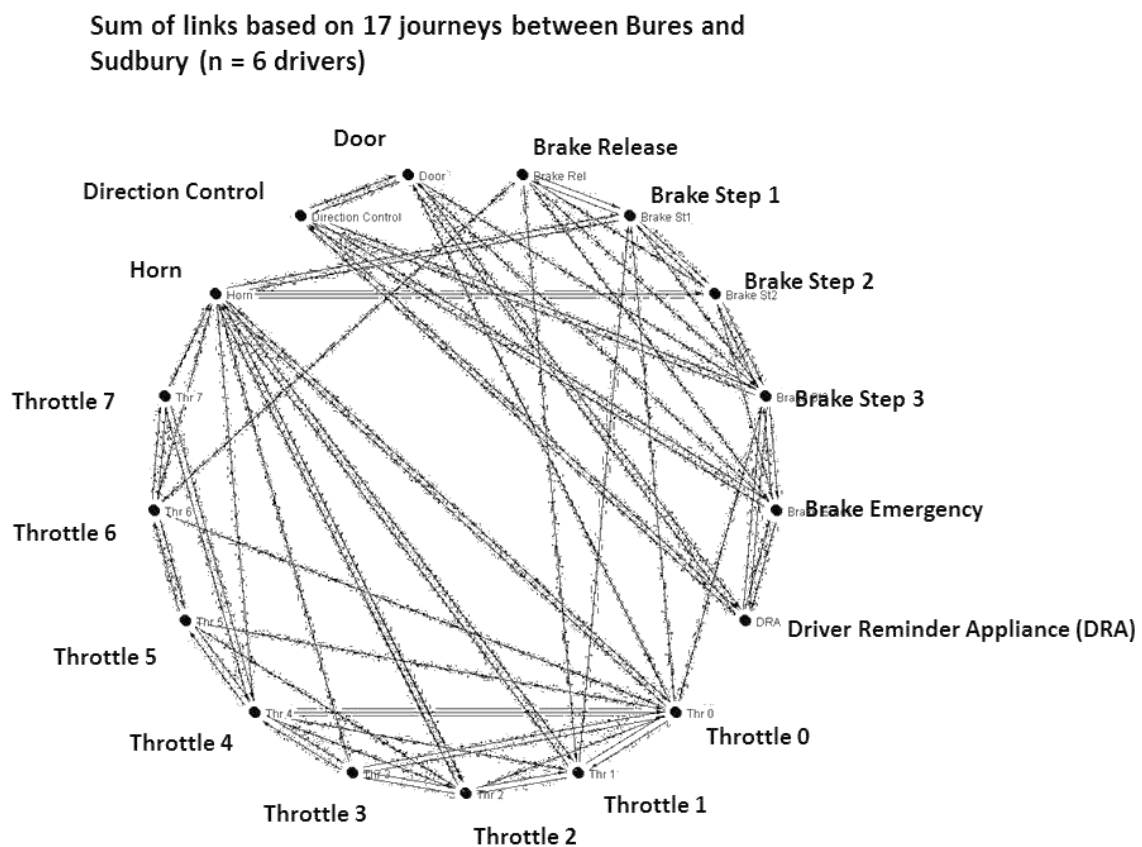


Figure 5 – Diagram illustrating the application of link analysis

Outputs

The pattern of control activities across all 17 journeys is summarised in the network diagram in Figure 6. This gives a visual representation of the pattern of connections that were

produced across all drivers and journeys in the analysis. The link diagram shows that some components of the cab interface are more heavily interconnected than others (as represented by thicker connecting lines), that there is an 'overall' level of connectivity, and that the number of links one needs to traverse to reach different pairs of nodes (controls) also differs.



Network Descriptives:

Number of nodes = 17

Number of links = 30

Figure 6 – Overall network diagram

To investigate whether individual drivers differ in their interactions with the cab interface and controls Graph Theory can be applied to the link matrices. An example is the network metric centrality, given by the formula:

$$Centrality = \sum_{j=1}^Y C_D(y^*) - C_D(y_j)$$

where the node with the greatest number of links converging upon it ($C_D(y^*)$), that which has the highest 'degree', is used to derive centrality values for every other node in the network. Simply stated, centrality is a way to identify which part of the interface is more heavily connected, and therefore more prominent in the network, compared to others. In practice, heavily connected nodes are more 'visible' to other nodes, serve as a conduit for control sequences more than other nodes, and will be subject to network influences more quickly and more often than other nodes.

The results of an analysis of centrality performed on the six drivers responsible for 17 journeys between Bures and Sudbury shows there is a distinct pattern identifying the importance of a control based on the number of times it is linked in sequence to another control. In this case, for many drivers, the lower positions of the brake and power controllers are dominant. Indeed, there is close concordance between all drivers in terms of the power controller, with progressively less emphasis given to higher power settings. There is much greater variation in terms of brake controller usage and three distinct 'signatures' seem to emerge. The first is the 'mean signature', which averages across all 17 journeys, and the one where equal and low importance is ascribed to all of the 'running' brake control positions (Steps 1 to 3 on the controller). Then there is another group of drivers who ascribe descending importance to Steps 1 to 3. The pattern for a third driver stands out from the group in that their responses cluster greatly around brake Step 2. Distinct patterns such as these grant access to leading indicators around driver strategy and technique.

What are the leading indicators?

Leading Indicator 6: Link analysis enables network diagrams to be created directly from recorder data. These represent human interactions with control interfaces in a graphical format, showing link frequency, direction and strength. Patterns of behaviour can be readily detected from these representations and potentially serve as leading indicators of driving technique.

Leading Indicator 7: The results of Link Analyses can be expressed as single numbers through the use of Graph Theory. Network metrics summarise different facets of human interaction with in cab/cockpit controls, such as response diversity, sequencing and clustering, and provide key performance indicators of when these interactions change.

Type of Risk Detected: Getting out of sequence because of task interruptions, unusual future responses being required, removal of normal environmental cues that trigger habitual behaviour, and deviations from well-practised routines: all of which are detectable via link diagrams and associated metrics. Loss of situational awareness due to missing information/links between elements of the system. Faulty allocation of attention due to experience/training related factors as revealed by driving strategies and signatures. Prior experience that overrides correct action because of high routine and minimal (new) inputs, changes in system mode meaning that a routine cluster of links now causes a different outcome, and conflicting information/links resulting in situations being misclassified.

Sensitivity and Decision Bias

Brief Description

Drivers have to respond to a wide range of stimuli in their environment, all of which is subject to some level of uncertainty. Tasks like these are not merely perceptual ones of seeing or hearing something, they are also cognitive: driver's and pilot's not only have to discriminate a 'stimulus' from within a 'noisy' environment, but correctly classify it and respond. Signal detection Theory (SDT) formalises these concepts by separating out a person's sensitivity to a stimuli (how easy it is to detect something) and their response bias (their preference for responding one way or another to the stimuli; Green & Swets, 1966). SDT helps us to understand why a particular 'stimulus', which might be very loud, visible or unambiguous, is not always responded to in the ways we expect (or vice versa). Signal Detection Theory classifies human responses to stimuli in the environment in four ways, either as a 'hit', 'miss', 'false alarm' or 'correct rejection'. The ability to accurately detect stimuli in the environment and correctly classify them is the desired outcome. Taking the example of AWS and the need to press a cancellation button in response to an in-cab warning, if the button was pressed in response to ANY warning indication this will ensure a 100% Hit rate but will also increase the rate of False Alarms. Accuracy in this case is low. If, on the other hand, the driver is trying to do the opposite, to avoid False Alarms and instead maximise Correct Rejections, they would not respond to ambiguous 'signals'. This would increase the number of Correct Rejections but it would also increase the number of 'Misses'. Accuracy in this case is also low. Signal Detection Theory enables us to separate sensitivity (d') from decision bias (C). Sensitivity is a measure of accuracy and tells us how easy it is to distinguish a particular environmental stimuli (e.g. an in cab warning). Decision bias tells us whether one response is more probable than another due to habit or other contextual influences.

Method

The proof of concept was tested using three railway journeys from Ipswich to Felixstowe.

Signal Detection Theory (SDT) was applied to the activation of the in-cab Automatic Warning System (AWS) in which driver responses could be categorised as follows:

Table 6 – Driver responses to the in-cab Automatic Warning System (AWS) organised into Signal Detection Theory categories

Status of AWS/TPWS System	Number	%	Categorisation
AWS Horn followed by reset	21	30%	Hit
AWS Horn followed by no response	0	0%	Miss
AWS Bell followed by reset or no activation followed by reset	22	31%	False Alarm
AWS Bell followed by no response	28	39%	Correct Rejection
Totals	71	100%	

Outputs

Sensitivity to a stimulus is given by the metric d-prime, which was calculated as follows:

$$d' = z(\text{FA}) - z(\text{H})$$

where $z(\text{H})$ is the number of Hits expressed as a z-value subtracted from the same Z-transformed False Alarm rate. The results obtained are shown in Table 7:

Table 7 – Drivers sensitivity to Automatic Warning System (AWS) alerts

Journey	Hits	Misses	Correct Rejections	False Alarms	d-prime

1	5	0	11	0	4.65
2	9	0	6	7	2.30
3	7	0	11	15	2.13
Mean					3.03

The d-prime figure measures the strength of the stimulus, which in this case is the in-cab warning provided by the AWS system. A value of 3.03 indicates drivers are highly sensitive to it: in this situation it is unambiguous and easy to discriminate from the wider background of noise, distractions, other contextual factors etc. Expressed more formally, the responses drivers' are providing when an AWS warning is overlain on top of the 'contextual noise' is 3.03 standard deviations 'different' from the responses they give when the signal is absent (and only the 'contextual noise' is present). Sensitivity provides an important leading indicator concerning the discriminability of information needed for driver's to develop accurate situational awareness. This is because the same 'stimuli' may yield different levels of sensitivity depending on external/contextual factors. A warning that was not expected, ambiguous, not fully understood or masked, for example, may lower sensitivity despite the fact that it is the same 'objective' warning (in terms of prominence, sound intensity level etc.).

Decision bias/criterion is given by the metric c , which was calculated as follows:

$$c = -\frac{Z(H) + Z(FA)}{2}$$

The results obtained are shown in Table 8. Decision Bias is independent of sensitivity and relates not to the discriminability of the 'signal' but to the payoffs involved in making one response in favour of another. Thus, regardless of how easy it is to discriminate a stimulus a

counter intuitive response may still be favoured. This is because the consequences of False Alarms, Misses and Correct Rejections vary with the context. Psychological research shows that decision bias is more unstable and situationally dependent than sensitivity and therefore a potentially valuable Leading Indicator.

Table 8 – Results of decision bias (c)

Journey	Hits	Misses	Correct Rejections	False Alarms	Criterion
1	5	0	11	0	0
2	9	0	6	7	-1.21
3	7	0	11	15	-1.26
				Mean	-1.24

The mean decision bias value across the three sampled drivers was $c = -1.24$ which indicates a liberal bias. Driver's make more responses which indicate the AWS signal is present than it is absent. In other words, they are prioritising False Alarms over Correct Rejections which, in turn, provides a clue as to the sorts of error that may be more likely to occur in future (i.e. warnings that are cancelled incorrectly). Assuming drivers' 'internal responses' to the AWS warning are normally distributed (as per Signal Detection Theory) it is possible to plot individual driver decision bias' into Figure 7 which, in turn, provides an important diagnostic tool in defining risky psychological/decision making states.

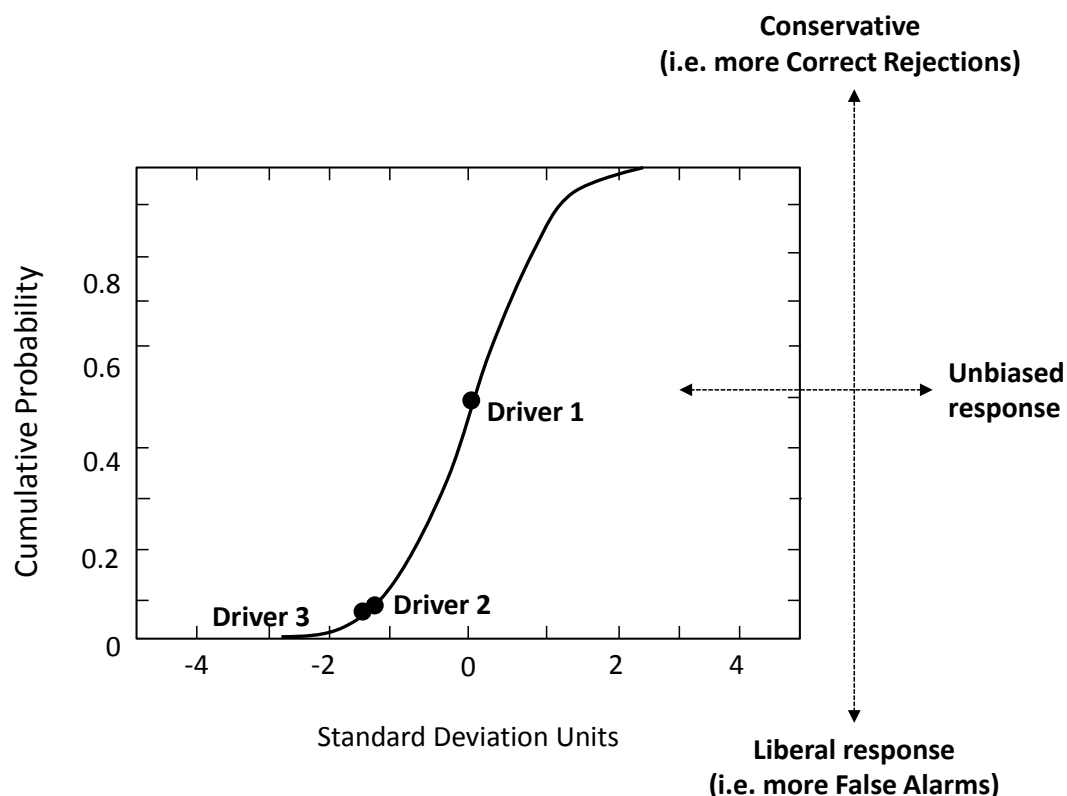


Figure 7 - Cumulative probability plot showing the extent and direction of response biases exhibited by the drivers.

According to Figure 7, Driver 1 shows no systematic bias in their responses to the AWS warning. They respond correctly to the AWS warning on every occasion and his/her False Alarm rate is zero. Drivers 2 and 3 are different. They are exhibiting a strong 'liberal response bias' meaning they are much more inclined to exhibit 'false alarm' responses (and behavioural clusters 2 and 3 above). With the ability to detect these changes in decision bias comes the possibility to analyse a) the extent to which different biases interact with accident/incident rates (i.e. is a liberal bias of this magnitude associated with particular types of risk) and b) how the context influences human decision making (and therefore how that context can be modified to 'un-bias' human responses).

What is the leading indicator?

Leading Indicator 8: Sensitivity provides a measure of how much useful information there is in the environment and the extent to which drivers can discriminate it from the background of contextual noise. Warnings, stimuli and so forth may, in an engineering sense, appear to be unambiguous yet they may be considerably less so cognitively. Sensitivity provides a measure of this which can, in turn, be associated with changing risk.

Leading Indicator 9: Decision bias reveals the likelihood that one type of driver response will be favoured and, in a wider application, how this interacts with risk. In a wider application it would be possible to examine decision bias in a systematic way looking at differences between drivers and between particular routes. This could provide insight into driving styles and indicate whether particular aspects of a route result in a shift in decision bias. For example, a specific AWS signal on a particular route may result in a high level of predictive cancellations/button pressing (high false alarms) relative to most others, identifying this as a more risky section of journey. Relationships such as these would need to be established based on a future research but the feature itself is now detectable.

Type of Risk Detected: Getting out of sequence because of changing levels of sensitivity to environmental cues that trigger habitual behaviour. Loss of situational awareness because of the detectability (or absence of detectability) of information in the environment. Risk of a failure to understand what environmental cues might mean due to high levels of decision bias. Faulty allocation of attention due to misdirected attention lowering sensitivity to other environmental stimuli. Prior experience overrides correct action as revealed by shifts in decision bias as a result of highly routine journeys, experience and habit, biases that favour one response type in a multi-modal system, the role of conflicting information in reducing sensitivity and increasing misclassifications.

Conclusions

Exabytes of data are routinely and continuously collected from normal journeys by On-Train Data Recorders (OTDR), but currently not used in a systematic fashion. At the same time, not only are the opportunities to use OTDR data after accidents diminishing because of improving safety trends but we are left with a class of Human Factors problem that is difficult to predict based on previous occurrences. Human Factors issues are a strategic risk and innovative new ways to work at this interface are required. The research reported in this paper advances this agenda. Table 9 summarises the nine leading indicator candidates that arise from coupling big-data to Human Factors methods and how they map to key risk types.

Table 9 – Summary table of leading indicators mapped to risk types

	Type of Risk Detected					
	Getting out of sequence	Loss of situational awareness	Faulty allocation of attention	Prior experience overrides correct action		
Workload Profile: high/low events						
Workload Profile: change						
Workload Profile: rate of change						
Process Charts: operations loading						
Process Charts: behavioural clusters						
Link Analysis: patterns of interaction						
Link Analysis: network metrics						
Sensitivity						
Decision bias						

The advantage of this approach is that Human Factors methods are able to bring with them a substantial and robust form of construct validity. They have already been shown to provide access to the psychological issues that are of interest in detecting key risks, and as such provide a novel framework for interrogating big-data from OTDR devices. Importantly, the process of review and proof of concept testing establishes the methods' scalability:

Human Factors methods 'can' be driven from this new type of data and, furthermore, are amenable to software-based implementation. This finding hints at future reductions in the analytical overhead required to perform these analyses, indeed, with more automation Human Factors leading indicators of the sort tested in this paper could become continuous key performance indicators, relying more on real-time data streams and analysis. The possibilities are tantalising, but having proved the concept a number of important future research tasks are required. Human Factors methods are able to demonstrate good construct validity and reliability (based on their substantial legacy of prior use and development) but another source of big-data needs to be combined with these candidate leading indicators. This is currently underway. This critical step is required in order to answer questions about whether identified risks around vigilance decrements, clusters of behaviour or decision bias find themselves implicated in actual incidents such as SPADs, overspeeding, wrong-side door releases and station over-runs. The sensitivity of the leading indicators needs to be established, along with their predictive validity in terms of actual risk outcomes. If this loop can be closed then we have a powerful new approach. The big-data 'haystack' may not be any smaller than it was previously, but the number of useful 'needles' to be found has been increased, and they are now easier to find.

Acknowledgements

This work was funded by the UK Economic and Physical Sciences Research Council (EPSRC) under Grant Reference: EPSRC EP/I036222/1. The involvement and support of the UK Rail Safety and Standards Board (RSSB), the Association of Train Operating Companies (ATOC) and Greater Anglia Trains is gratefully acknowledged.

References

- Bartlett, F.C. (1932). Remembering: A Study in Experimental and Social Psychology. Cambridge: Cambridge University Press.
- Cullen, W.D. (2001). The Ladbroke Grove Rail Inquiry: Parts 1 and 2 Report. London, UK: HSE Books.
- Dismukes, R.K., (2007). Prospective Memory in Aviation and Everyday Settings, In: Kliegel, M., McDaniel, M.A. & Einstein, G.O. (Eds.), Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives. Mahwah: Erlbaum.
- Endsley, M. R. (2012). Designing for situation awareness: an approach to user-centred design. Boca Raton, FL, CRC Press.
- European Transport Safety Council. (2003). Transport safety performance in the EU: A statistical overview. Brussels: European Transport Safety Council. Retrieved July 17, 2012, from: <http://www.etsc.be/documents/statoverv.pdf>. Ferris, Sarter & Wickens, 2010;
- Evans, A.W. (2011). Fatal train accidents on Europe's railways: 1980-2009. Accident Analysis and Prevention, 43(1), 391-401.
- Eysenck, M.W. (1982). Attention and arousal. Cognition and performance: Berlin:
- Geisler, S., Quiz, C., Schiffer, S. & Jarke, M. (2012). An evaluation framework for traffic information systems based on data streams. Transportation Research Part C, 23, 29-55.
- Gilbreth, F. B. & Gilbreth, L. M. (1921). Industrial efficiency. New York: American Society of Mechanical Engineers.
- Gobet, F. (1998). Expert memory: A comparison of four theories. Cognition , 66, 115-152
- Graf, P. (2011). Special issue editorial: prospective memory in 2010. Canadian Journal of Experimental Psychology; Revue Canadienne de Psychologie Experimentale, 65(1), 1-2.

Walker, G. H. & Strathie, A. (2016). *Big data and ergonomics methods: a new paradigm for tackling strategic transport safety risks*. *Applied Ergonomics*, 53, 298-311

Green, D.M., Swets J.A. (1966). Signal Detection Theory and Psychophysics. New York: Wiley

Groeger, J. A., Clegg, B. A., & O'Shea, G. (2005). Conjunction in simulated railway signals: A cautionary note. Applied Cognitive Psychology. 19(8), 973-984.

Hart, C. (2003). Global Aviation Information Network. in Proceedings of the Human Factors Workshop: Improving Railroad Safety Through Understanding Close Calls. Baltimore, MD.

Hicks, J.L., Marsh, R.L., & Cook, G.I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. Journal of Memory and Language, 53, 430–444

Karwowski, W. (2001). International encyclopaedia of ergonomics and human factors. Boca-Raton, FL: CRC Press.

Laney, D. (2001). 3D data management: controlling data volume, velocity and variety. Meta Group.

Loft, S., Smith, R. E., & Bhaskara, A. (2011). Prospective memory in an air traffic control simulation: External aids that signal when to act. Journal of Experimental Psychology: Applied, 17, 60-70.

Mackworth, N. H..(1948). The breakdown of vigilance during prolonged visual search. Quarterly Journal of Experimental Psychology, 1, 6–21.

McLeod, R. W., Walker, G. H., & Moray, N. (2005). Analysing and modelling train driver performance. Applied Ergonomics, 36, 671-680.

McMains, S.A., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. Journal of Neuroscience 31(2), 587–597.

Walker, G. H. & Strathie, A. (2016). *Big data and ergonomics methods: a new paradigm for tackling strategic transport safety risks*. *Applied Ergonomics*, 53, 298-311

Meacham, J. A., & Leiman, B. (1982) Remembering to perform future actions. In: Neisser, U. (Ed.), Memory observed: Remembering in natural contexts. San Francisco, CA: W. H. Freeman and Company, 327–336.

Mitchell, W.C., & Burns, A.F. (1938). Statistical indicators of cyclical revivals, NBER Bulletin 69, New York; reprinted as Chapter 6 of G.H. Moore (ed.) Business Cycle Indicators, Princeton, NJ, Princeton University Press, 1961.

Morcom, A.R. (1970). Flight Data Recording Systems: A brief survey of the past developments, current status and future trends in flight recording for accident investigation and operational purposes. Aircraft Engineering and Aerospace Technology, 42 (7)12 – 16.

Neisser, U. (1979). The control of information pickup in selective looking. In: A.D. Pick (Ed.), Perception and its development: A tribute to Eleanor J. Gibson, 201–219. Hillsdale, NJ: Erlbaum.

Norman, D.A. (1981). Categorization of action slips. Psychological Review, 88(1), 1-15.

Office of Rail Regulation (ORR) (2012), National rail trends yearbook, 2010 -2011. London: ORR.

Park, D.C., Hertzog, C., Kidder, D.P., Morrell, R.W., & Mayhorn, C.B. (1997). The effect of age on event-based and time-based prospective memory. Psychology and Aging, 12(2), 314-327.

Rail Safety & Standards Board (RSSB) (2009). The railway strategic safety plan 2009 – 2014. London: Rail Safety and Standards Board.

Rogers, A., Evans, R., and Wright, M. (2009). Leading indicators for assessing reduction in risk of long latency diseases. Greenstreet Berman Limited for Health and Safety Executive.

Walker, G. H. & Strathie, A. (2016). *Big data and ergonomics methods: a new paradigm for tackling strategic transport safety risks*. *Applied Ergonomics*, 53, 298-311

Salmon, P. M., Stanton, N. A., Walker, G. H., & Jenkins, D. P. (2009). Distributed situation awareness: advances in theory, measurement and application to teamwork. Farnham, UK: Ashgate.

Simons, D.J. & Chabris, C.F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28, 1059-1074.

Springer-Verlag; 1982.Walker, G. H. & Strathie, A. (2012). EPSRC EP/IO36222/1 Flight Data Monitoring, On Train Data Recording and Human Factors: Review of the Knowledge Base. Heriot-Watt University, Edinburgh. Unpublished Report.

Stanton, N. A., Young, M. S. & McCaulder, B. (1997). Drive-by-wire: the case of driver workload and reclaiming control with adaptive cruise control. *Safety Science*, 27(2-3), 149-159.

Stanton, N. A. & Young, M. S. (1999). A guide to methodology in ergonomics: designing for human use. London: Taylor and Francis.

Stanton, N. A. & Walker, G. H. (2011). Exploring the psychological factors involved in the Ladbroke Grove rail accident. *Accident Analysis and Prevention*, 43(3), 1117 - 1127.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C. & Jenkins, D. P. (2013). Human factors methods: a practical guide for engineering and design. Farnham: Ashgate.

Stanton, N.A, Chambers, P.R.G. & Piggott, J. (2001) Situational awareness and safety. *Safety Science*, 39, 189-204

Step Change in Safety (2003). Leading Performance Indicators Guidance for Effective Use Brochure. Aberdeen, UK: Step Change in Safety.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, 49, 83-90.

Walker, G. H. & Strathie, A. (2016). *Big data and ergonomics methods: a new paradigm for tackling strategic transport safety risks*. *Applied Ergonomics*, 53, 298-311

Uff, J. (2000). The Southall rail accident inquiry report. London: HMSO.

Wolmar (2012, May) Rail Magazine, 697: safety is the hidden success story

<http://www.christianwolmar.co.uk/2012/05/rail-697-safety-is-the-hidden-success-story>

Wu, X. & Liu, H. X. (2014). Using high-resolution event-based data for traffic modelling and control: an overview. Transportation Research Part C, 42, 28-43.

Young, M. S. & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. Human Factors, 44(3), 365-375.